Volume 237, number 1,2, 178–182 FEB 06302 September 1988

Amino acid sequence microheterogeneities of a type I cytokeratin of M_r 51000 from *Xenopus laevis* epidermis

Werner Hoffmann, Sylvia Sterrer and Andreas Königstorfer

Max-Planck-Institut für Psychiatrie, Abteilung Neurochemie, Martinsried, FRG and Institut für Molekularbiologie der Österreichischen Akademie der Wissenschaften, Salzburg, Austria

Received 27 July 1988

The complete cDNA-derived sequence of a type I cytokeratin (designated no. 3) from *Xenopus laevis* skin is described. The deduced protein has an M_r of 51888 and consists of a glycine-rich head domain, a well-conserved α-helical region and a tail rich in hydroxyamino acid residues. Various cDNA clones encoding two different mRNAs were isolated that differed by short deletions/insertions and point mutations. These microheterogeneities are mainly located in a 'hypervariable region' at the C-terminal non-α-helical region.

Cytokeratin; Intermediate filament; cDNA cloning; (*Xenopus*, Frog skin)

1. INTRODUCTION

Like desmin, vimentin, glial filament proteins and neurofilaments, the cytokeratins form one of the five major types of intermediate filaments (IF) with a diameter of 7–11 nm [1]. This very complex family of keratin-type proteins is characteristic for epithelial cells and in human tissues 19 different polypeptides have been identified [2].

In *Xenopus laevis* the keratin pattern reveals drastic developmental changes during epidermal maturation [3]. However, in adult *X. laevis* epidermis, a unique set of cytokeratins is synthesized ranging from M_r 49000 to 64000 [4,5]. These polypeptides can be divided into two different classes: type II molecules are basic with M_r 64000 and are closely related to each other (designated nos 1a–c). On the other hand, acidic type I molecules are more diverse, and this class consists of at least three different polypeptides with M_r of 53000 (no. 2), 51000 (no. 3) and 49000 (no. 4). Using in vitro translation of hybrid-selected mRNAs, partial cDNA clones encoding cytokeratin nos 1a–c and cytokeratin no. 3 could be identified [6,7]. These cytokeratins (M_r 64000 and 51000) are not expressed in oocytes, unfertilized eggs, liver or skeletal muscles of adult frogs [8]. During development, the mRNAs for the 64 kDa polypeptides first appear at stages 48–52, and that for the 51 kDa protein before stage 46 [9]. Here we present the full-length sequence of the type I cytokeratin, with M_r 51000 as deduced from cDNA cloning.

Correspondence (present) address: W. Hoffmann, Max-Planck-Institut für Psychiatrie, Abteilung Neurochemie, D-8033 Martinsried, FRG

The nucleotide sequence presented here has been submitted to the EMBL/GenBank database under the accession no. Y00968

2. MATERIALS AND METHODS

All manipulations with DNA including cDNA cloning (pUF < 1000 clones) were performed as described [10]. Additionally, for the production of full-length clones, new improved cDNA libraries have been constructed (pUF > 1000 clones) as reported in [11]. Screening of cDNA libraries with synthetic oligodeoxynucleotides has been described elsewhere [12] with the exception that Hybond-N membranes (Amersham) were used. DNA sequences were determined by a combination of both the Maxam-Gilbert protocol [13] using a modified G + A reaction [14] and the dideoxy method [15] after subcloning into different M13 vectors [16]. DNA sequencing data were computerized using programs purchased from Teufl-Software (Salzburg). A commercial program from the University of Wisconsin Genetics Computer Group (UWGCG, version 5.1) was used for the prediction of protein secondary structures.

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50

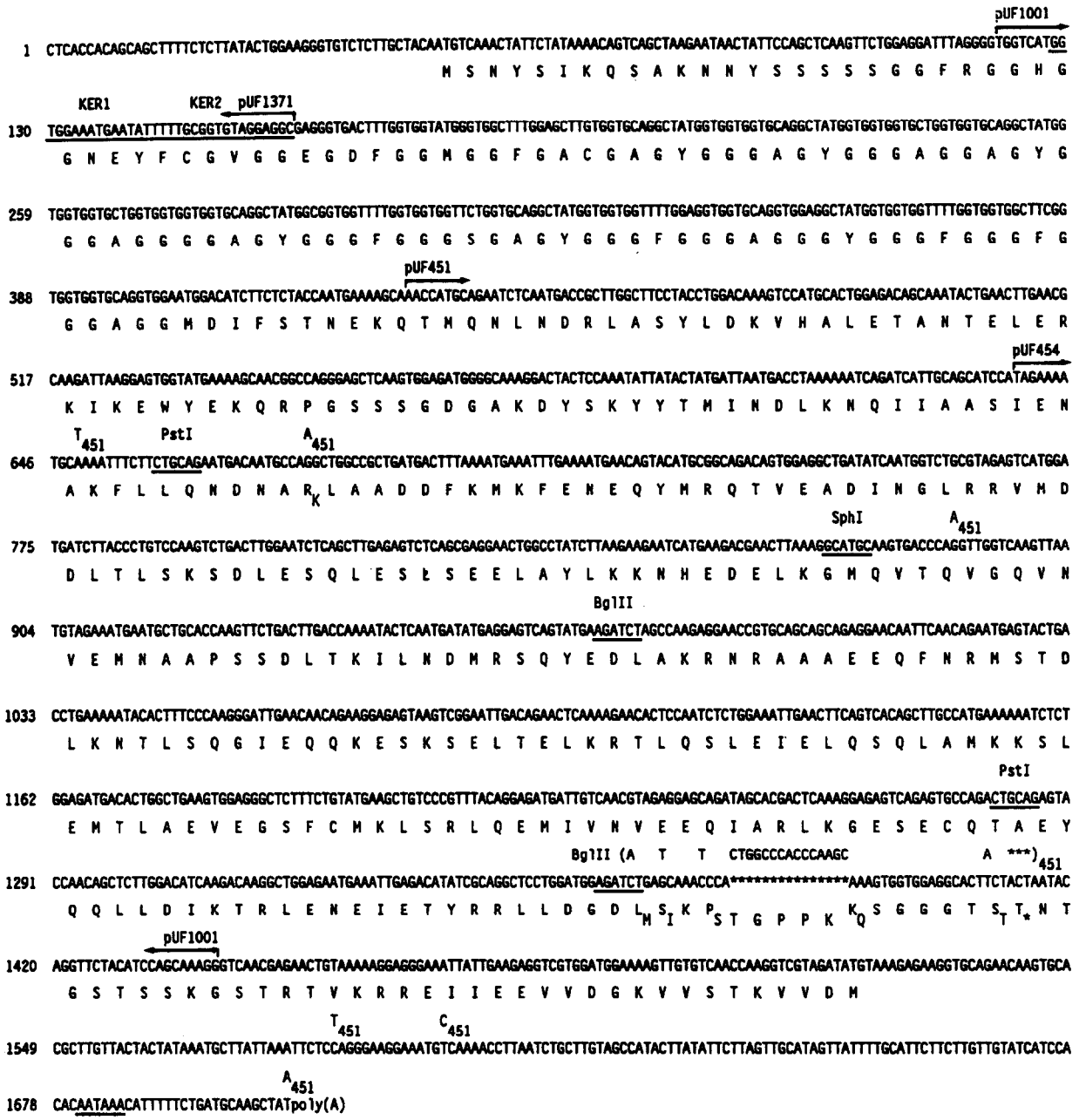

Fig.1. Nucleotide sequence and deduced amino acid sequence of XL51. The full-length sequence was obtained from cDNA clones pUF1371, pUF1001 and pUF454. In the line above, the differences observed in cDNA clone pUF451 are displayed, when compared with the combined full-length sequence. Asterisks were introduced to maximize homology (introducing gaps into the sequences). Restriction sites, the polyadenylation signal AATAAA, and the positions of the synthetic oligonucleotides KER1 and KER2 are underlined. The lower line of the amino acid sequence represents the differences observed in pUF451 when compared with the full-length sequence deduced from pUF1371, pUF1001 and pUF454.

3. RESULTS

Starting with the *Bgl*II-*Eco*RI fragment of cDNA clone pUF29 [7], which encodes only a C-terminal portion of the type I cytokeratin with M_r 51 000 (designated no. 3), we screened the original library of *X. laevis* skin [10] for longer clones and obtained pUF451 and pUF454 (fig.1). Subsequently, the smallest *Pst*I fragment of pUF451 (corresponding to the most 5′-sequences obtained so far) was used as the next hybridization probe for an improved library resulting in the isolation of clone pUF1001. In order to obtain the 5′-end of the mRNA cloned, a different approach was applied: based on the 5′-sequence of clone pUF1001 the oligonucleotides KER1 d(AATATTCATTTCCACC) and KER2 d(GCCTCCTACACCGCAAAA) were constructed. KER2 was used as a primer to produce a partially enriched cDNA library of *X. laevis* skin, which was screened with KER1. Of about 500 clones, a single positive designated pUF1371 could be picked.

4. DISCUSSION

4.1. *Nucleic acid sequence*

Fig.1 represents the combined full length sequence of a type I cytokeratin (XL51) as deduced from cDNA-clones pUF1371, pUF1001 and pUF454. The overlapping sequences of all three clones are identical (40 and 787 nucleotides, respectively). This combined sequence differs in a few point mutations and deletions/insertions from the sequence derived from the insert of pUF451. Both new sequences presented here also differ from the previously published sequence of pUF29 [7]. These differences are exclusively located at the extreme 3′-non-coding-region. Due to processing errors, the sequence of pUF29 [7] had to be corrected at four different positions (position 42, C to G; 292, A to G; 507, T to C, and insertion of T after 596 according to the numbering in [7].

The microheterogeneities observed probably arise from the quasi-tetrapolid genome of *X. laevis* [17] resulting from a duplication of the genome about 30 million years ago [18]. This long time period could easily account for the evolutionary drift of these two sets of genes. Similar microheterogeneities have previously been reported repeatedly, e.g. for vitellogenin [19].

The length of the full sequence presented here (1692 nucleotides without the poly(A) tail) fits very well with the result from the Northern analysis showing a band of about 1.8 kilobases [7]. The deduced full-length amino acid sequence (derived from pUF1371, pUF1001 and pUF454) results in

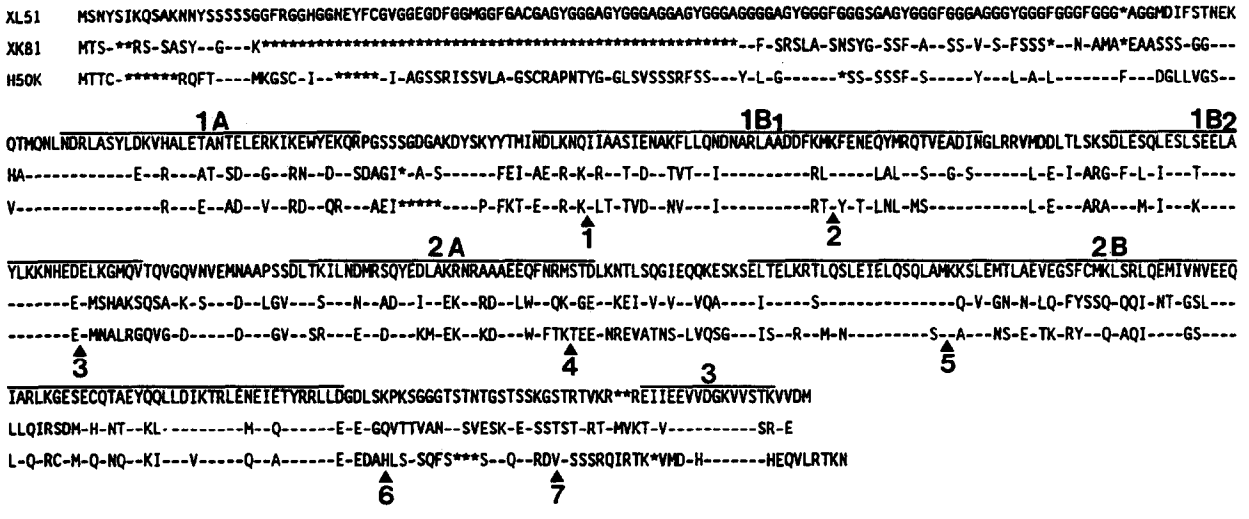

Fig.2. Comparison of XL51 (see fig.1) with cytokeratin XK81 present in gastrulae of *X. laevis* [26] and human type I cytokeratin H50K [21,25]. Asterisks were introduced to maximize homology (introducing gaps into the sequences) and bars in XK81 and H50K represent identical amino acid residues when compared with XL51. Also shown are the predicted α-helical regions (1A, 1B1, 1B2, 2A, 2B, 3) according to Garnier et al. [24]. Black triangles identify positions of the seven introns as determined for H50K [25].

an M_r of 51 888 including the initiating methionine, which is most likely lost in the mature protein. This is in very good agreement with the M_r determined by gel electrophoresis, which is 51 000 [7]. The 3'-non-coding-region of the mRNA presented in fig.1 comprises 185 nucleotides not including the poly(A) tail. This is considerably shorter than that of an mRNA encoding a type II cytokeratin of *X. laevis* epidermis (M_r 64 000), which consists of an 3'-untranslated region of 476 nucleotides [20]. A similar result has been reported for human cytokeratins [21,22].

Remarkable is the high similarity of about 28 bases at the extreme 3'-ends of the mRNA encoding a type I cytokeratin (M_r 59 000) of mouse [23] and of the sequence presented in fig.1.

4.2. *Secondary structure*

A secondary structure prediction according to Garnier et al. [24] revealed three major domains (see fig.2). The glycine-rich head domain at the N-terminal has mainly β-turn structure and in the middle of the molecule a cluster of helical regions can be distinguished (coils 1A, 1B1, 1B2, 2A, 2B). The tail is rich in hydroxyamino acid residues also indicating β-turns, and a short α-helical region (coil 3) can be predicted only at the extreme C-terminal.

Nearly all microheterogeneities occur within the C-terminal tail region. The large insertion of 15 nucleotides in pUF451 mainly affects the position of proline residues, marking the end of coil 2B. This hypervariable region is located just at the end of the coiled-coil structure, similarly to that previously described for various basic cytokeratins of *X. laevis* [6]. Based on the homology with human 50 kDa cytokeratin [21,25] (see fig.2), this hypervariable region would be entirely located within exon 7.

4.3. *Homologies*

As illustrated in fig.2, the sequence presented here (XL51) shows strong homology with other type I cytokeratins from frog (XK81) and man (H50K), which is no. 14 according to the catalog of human cytokeratins [2]. Clearly, the highest degree of homology is concentrated within the α-helical rod domain in the middle. The non-α-helical head and tail regions of the molecules show only the well-conserved potential to form β-turns. At the extreme C-terminal end there is again a short but highly conserved stretch of about 11 amino acid residues, which can form an α-helical structure (coil 3).

The sequence comparison in fig.2 also displays clearly that XK81, present mainly during gastrulation but also present in a few organs of adult *X. laevis* (Franke, W.W., personal communication), shares more similarity with human cytokeratin no. 14 in the α-helical domain than does XL51. Conversely, the human sequence is closer to XL51 than XK81 with respect to the length of the non-α-helical head and tail domains, accounting for the very similar molecular masses of H50K and XL51.

In contrast to all other type I cytokeratins sequenced, the non-α-helical head domain of XL51 contains three acidic residues at positions 30, 38 and 40 (see fig.2). Typically, this domain contains a few basic residues.

In conclusion, we believe that the type I cytokeratin (M_r 51 000) presented here is probably not the direct amphibian counterpart of human cytokeratin no. 14, despite their very similar M_r values. It might, however, be the equivalent of no. 13 (M_r 54 000) or the more acidic no. 15 (M_r 50 000) of the human catalog [2].

Acknowledgements: We thank Eva-Maria Gertzen (Martinsried) for valuable technical assistance and U. Schimanko and D. Weigand (Martinsried) for quick support with synthetic oligonucleotides. We are also indebted to Dr P. Küster (Martinsried) for help in handling the computer and Dr D. Edgar (Martinsried) for critical reading of the manuscript. Part of this project was supported by a grant from the Bundesministerium für Wissenschaft und Forschung (Austria).

REFERENCES

[1] Steinert, P.M. (1985) Annu. Rev. Cell Biol. 1, 41–65.
[2] Moll, R., Franke, W.W., Schiller, D.L., Geiger, B. and Krepler, R. (1982) Cell 31, 11–24.
[3] Ellison, T.R., Mathisen, P.M. and Miller, L. (1985) Dev. Biol. 112, 329–337.
[4] Franz, J.K., Gall, L., Williams, M.A., Picheral, B. and Franke, W.W. (1983) Proc. Natl. Acad. Sci. USA 80, 6254–6258.
[5] Franz, J.K. (1982) Diploma Thesis, Faculty of Biology, University of Heidelberg, pp. 1–110.
[6] Hoffmann, W., Franz, J.K. and Franke, W.W. (1985) J. Mol. Biol. 184, 713–724.
[7] Hoffmann, W. and Franz, J.K. (1984) EMBO J. 3, 1301–1306.
[8] Franz, J.K. and Franke, W.W. (1986) Proc. Natl. Acad. Sci. USA 83, 6475–6479.

[9] Mathisen, P.M. and Miller, L. (1987) Genes Devel. 1, 1107-1117.
[10] Hoffmann, W., Bach, T.C., Seliger, H. and Kreil, G. (1983) EMBO J. 2, 111-114.
[11] Hoffmann, W. (1988) J. Biol. Chem. 263, 7686-7690.
[12] Singer-Sam, J., Simmer, R.L., Keith, D.H., Shively, L., Teplitz, M., Itakura, K., Gartler, S.M. and Riggs, A.D. (1983) Proc. Natl. Acad. Sci. USA 80, 802-806.
[13] Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
[14] Cooke, N.E., Coit, D., Shine, J., Baxter, J.D. and Martial, J.A. (1981) J. Biol. Chem. 256, 4007-4016.
[15] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
[16] Messing, J. (1983) Methods Enzymol. 101, 20-78.
[17] Bisbee, C.A., Baker, M.A., Wilson, A.C., Azimi, I.H. and Fischberg, M. (1977) Science 195, 785-787.
[18] Thiébaud, C.H. and Fischberg, M. (1977) Chromosoma 59, 253-257.
[19] Wahli, W. and Dawid, I.B. (1979) Cell 16, 535-549.
[20] Königstorfer, A. (1987) Master Thesis, Faculty of Science, University of Salzburg, pp. 1-25.
[21] Hanukoglu, I. and Fuchs, E. (1982) Cell 31, 243-252.
[22] Hanukoglu, I. and Fuchs, E. (1983) Cell 33, 915-924.
[23] Krieg, T.M., Schafer, M.P., Cheng, C.K., Filpula, D., Flaherty, P., Steinert, P. and Roop, D.R. (1985) J. Biol. Chem. 260, 5867-5870.
[24] Garnier, J., Osguthorpe, D.J. and Robson, B. (1978) J. Mol. Biol. 120, 97-120.
[25] Marchuk, D., McCrohon, S. and Fuchs, E. (1985) Proc. Natl. Acad. Sci. USA 82, 1609-1613.
[26] Jonas, E., Sargent, T.D. and Dawid, I.B. (1985) Proc. Natl. Acad. Sci. USA 82, 5413-5417.